HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2019 June 14.

Published in final edited form as:

Nat Genet. 2018 July ; 50(7): 928–936. doi:10.1038/s41588-018-0142-8.

Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci

A full list of authors and affiliations appears at the end of the article.

Abstract

Genome-wide association studies (GWAS) and fine-mapping efforts to date have identified more than 100 prostate cancer (PrCa)-susceptibility loci. We meta-analyzed genotype data from a custom high-density array of 46,939 PrCa cases and 27,910 controls of European ancestry with previously genotyped data of 32,255 PrCa cases and 33,202 controls of European ancestry. Our analysis identified 62 novel loci associated ($P<5.0\times10^{-8}$) with PrCa and one locus significantly associated with early-onset PrCa (≤55 years). Our findings include missense variants rs1800057 (odds ratio (OR) = 1.16; $P = 8.2\times10^{-9}$; G>C, p.Pro1054Arg) in *ATM* and rs2066827 (OR = 1.06; $P = 2.3 \times 10^{-9}$; T>G, p.Val109Gly) in *CDKN1B.* The combination of all loci captured 28.4% of the PrCa familial relative risk, and a polygenic risk score conferred an elevated PrCa risk for men in the ninetieth to ninety-ninth percentiles (relative risk = 2.69; 95% confidence interval (CI): 2.55–2.82) and first percentile (relative risk = 5.71; 95% CI: 5.04–6.48) risk stratum compared with the population average. These findings improve risk prediction, enhance fine-mapping, and provide insight into the underlying biology of PrCa[1].

Although PrCa is the most common noncutaneous cancer among men in the Western world, and one in seven men will be diagnosed during their lifetime[2], very few modifiable risk factors have been established[3]. Epidemiological studies have identified age, positive family history, and ancestry as the most prominent risk factors for PrCa[4–7]. PrCa incidence is

* frs2@case.edu; aa461@medschl.cam.ac.uk; ros.eeles@icr.ac.uk. **Correspondence and requests for materials** should be addressed to F.R.S. or A.A.A. or R.A.E.

Author contributions

Writing group: F.R.S., C.A.H., D.V.C., A.A.A.O., S.I.B., M. Ahmed, Z.S.K.-J., and R. A.E. Conception and coordination of OncoArray synthesis: F.R.S., C.I.A., D.F.E., S. J.C., C.A.H., B.E.H., and F.W. Database management: S.B., M.N.B., X.S., K.M., and A.L. Bioinformatics support: E.J.S., T.D., D. Leongamornlert, E.A., C.C.-B., and C.G. Genotyping calling and QC: L.F., J.D., and J.T. Provision of DNA samples and/or phenotypic data: V.L.S., S.M.G., B.D.C., C.M.T., P.J.G., I.M.T., J.B., S.C., L. Moya, J.C., L.H., W.T., G.P.R., H.G., M. Aly, T.N., P. Pharoah, N.P., J.S., T.L.J.T., C. Slavov, A.A., D.A., S.W., A.W., N.H., C.M.L.W., A.M.D., N.B., L.A.M., E.G., G.L.A., O.C.,G.C.T., S.K., L.E.B.F., K.D.S., T.F.O., M.B., L. Maehle, E.M.G., D.E.N., J.L.D., F.C.H., R.M.M., R.C.T., T.J.K., R.J.H., N.E.F., A.F., S.A.I., M.C. Stern, B.S.R., S.L.K., H.O., Y.-J.L., H.-W.Z., N.F., X.M., X.G., G.W., Z.S., G.G.G., M.C. Southey, R.J.M., L.M.F., A.S.K., B.M.K., J. Lubinski, G.C.-V., K.L.P., M.S., J.Y.P., T.A.S., H.-Y.L., J.L.S., C.C., D.W., J. Lim, E.A.O., M.S.G., B.G.N., S.F.N., M.W., R.B., M.A.R., P.I., H.B., K.C., B.H.,C.M., M.L., T.S., J.K., C.J.L., E.M.J., M.R.T., P. Paulo, M.C., S.L.N., L.S., Y.C.D., K.D.R., G.D.M., P.O., A.R., J. Llorca, S.-H.T., D.W.L., L.F.N., D. Lessel, M.G., T.K., R.K., N.U., S.S., C. Sipeky, V.M., M.P., F Canzian, S.J., T.V.d.B., S. Larkin, P.A.T., C.A.H., M.G.D., J.E.C., M.E.M., M.J.R., G.J., R.H.N.v.S., F.M., T.T., Y.A.K., J.X., K.-T.K., L.C.-A., H.P., A.M., S.N.T., S.K.M., D.J.S., S. Lindstrom, C.T., J.M., D.J.H., E.R., A.S., F Claessens, L.N.K., L.L.M., R.N.H., M.J.M., Z.C., P.K., F.W., S.J.C., B.E.H., C.A.-H., R.A.E., A.V., A.G.-C., B.F.D., G.C.-T., APCB investigators, IMPACT Study, Canary PASS investigators, BPC3, PRACTICAL, CAPS, PEGASUS, GAME-ON/ELLIPSE, and Profile Study Steering Committee. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0142-8.

highest among men of African ancestry, followed by men of European and Asian ancestries. These observations of ancestral differences in PrCa risk, in conjunction with studies demonstrating the influence of family history[8,9], highlight the contribution of genetics to PrCa etiology[10]. Our previous work, using a multiplicative model, has estimated that more than 1,800 common SNPs independently contribute to PrCa risk among populations of European ancestry[11]. GWAS have reported more than 100 of these PrCa variants across multiancestral populations, most of which were identified in populations of European ancestry[12–29].

To facilitate additional discovery of PrCa genetic risk factors, we developed a custom high-density genotyping array, the OncoArray, including a 260,000-SNP backbone designed to adequately tag most common genetic variants (minor allele frequency (MAF) >5% in Europeans), and 310,000 SNPs from meta-analyses of five cancers (breast, colorectal, lung, ovarian, and prostate)[30]. Approximately 80,000 PrCa-specific markers derived from our previous multiancestral meta-analysis[12] (including populations of European, African American, Japanese, and Latino ancestry), fine-mapping of known PrCa loci, and candidate SNPs nominated by study collaborators were included on the OncoArray. We assembled a new PrCa sample series from 52 studies to genotype with the OncoArray (Supplementary Tables 1 and 2). After application of rigorous quality control (QC) criteria and removal of overlapping samples from previous studies, our OncoArray sample yielded 46,939 PrCa cases and 27,910 controls without a known diagnosis of PrCa and of European ancestry for analysis (Methods and Supplementary Table 3). Genotypes were phased and imputed to the cosmopolitan panel of the 1000 Genomes Project (1KGP; June 2014 release) in SHAPEIT[31] and IMPUTEv2 (ref.[32]) software (Methods and Supplementary Table 3). We performed a fixed-effects meta-analysis combining the summary statistics from our OncoArray analysis and seven previous PrCa GWAS or high-density SNP panels of European ancestry imputed to the 1KGP The final meta-analysis included 79,194 PrCa cases and 61,112 controls without a known diagnosis of PrCa (Fig. 1).

We performed study- and consortia-specific meta-analyses to identify novel PrCa risk loci. We established a *P*-value threshold of 5.0×10^{-8} to determine genome-wide significance. Our large sample size enabled several stratified meta-analyses focusing on key clinical and biological parameters (Methods and Supplementary Tables 4 and 5). All analyses used a likelihood-ratio test to minimize bias from rare variants, and a logistic-regression framework was used for all analyses, except for the Gleason score, for which linear regression was used. The genotype doses were incorporated in an allelic genetic model. The average λ_{1000}, an inflation statistic calibrated to a sample size of 1,000 cases and 1,000 controls[33], across the eight GWAS studies was 1.02 (range 0.98–1.09) and 1.00 for the overall meta-analysis (Supplementary Table 6). Our novel findings excluded variants within defined fine-mapped regions of previously reported PrCa risk loci (Supplementary Table 7).

After the exclusion of all known susceptibility regions (fine-mapping coordinates provided in Supplementary Table 7 and Supplementary Note), we identified 64 loci associated with overall PrCa susceptibility and 1 locus associated with early-onset PrCa ($P < 5.0 \times 10^{-8}$) in the meta-analysis (Supplementary Fig. 1), of which 53 were imputed, and 12 were genotyped with the OncoArray. The cluster plots for the genotyped makers are presented in

Supplementary Fig. 2. Although most of the imputed markers were of high quality, with an average imputed r^2 >0.80 for 61 of the 65 loci across all contributing GWAS (Supplementary Table 8), we closely examined four variants with a poor imputation quality score (r^2 <0.80) in the OncoArray samples by inspecting linkage disequilibrium (LD) plots including only genotyped SNPs from the OncoArray and performing an imputation QC assessment (Methods). After reviewing the LD plots and the imputation QC, we determined that loci rs6602880 and rs144166867 were probably false positives due to imputation artifacts (Supplementary Fig. 3 and Supplementary Table 9). Overall, we identified 62 novel loci associated with overall PrCa risk and one novel locus associated with early-onset PrCa (Table 1). The consortia-specific associations were consistent across the eight contributing GWAS studies (Supplementary Table 10).

We performed several stratified analyses defined by clinical and population parameters. We detected a novel variant, rs138004030, which was significantly associated with early-onset disease (Table 1) but was only nominally significant for overall PrCa risk ($P = 0.02$). In addition, we detected four markers significantly associated ($P < 5 \times 10^{-8}$) with advanced PrCa and two markers associated with early-onset PrCa (Supplementary Table 11). However, the case-only analyses of these markers indicated marginal statistical significance ($P < 1.0 \times 10^{-3}$). Additionally, these markers were in LD with nearby index markers associated with overall PrCa and were not significantly associated with overall aggressive disease after adjustment for the index marker (Supplementary Table 11). A similar association pattern was observed for rs111599055, which was in LD with rs7295014 (r^2 = 0.54), a marker associated with overall disease. The early-onset marker rs77777548 was independent of novel and known PrCa-risk loci. However, the marker was relatively rare (effect-allele frequency <0.02), was indicated as monomorphic in the 1KGP, and had a moderate imputation quality score (average $r^2 = 0.57$); hence, we did not include it in further analyses.

Among the 63 novel associations, 38 variants were found to be located within gene-rich regions (Supplementary Table 12): intronic (32 SNPs), missense (4 SNPs), and 3′ untranslated region (UTR) (2 SNPs). Analyses of expression quantitative trait loci (eQTL) in The Cancer Genome Atlas (TCGA) database identified statistically significant associations ($P < 0.05$; Supplementary Table 12) in normal PrCa tissue for 17 of the novel associations, including both 3′-UTR SNPs and 11 of the 32 intronic SNPs. Cis-eQTL associations were identified for 3′-UTR variant rs1048169 with *HAUS6* (3′-UTR) and intronic variants rs182314334 with *MBNL1,* rs4976790 with *COL23A1,* rs9469899 with *UHRF1BP1,* rs878987 with *B3GAT1,* rs11629412 with *PAX9,* and rs11666569 with *MYO9B.* The eQTL associations were consistent with the observed PrCa-SNP associations, given that we assessed colocalization between the GWAS and eQTL SNPs. The TCGA data analysis did not identify an eQTL association with any of the four missense SNPs.

We assessed the association of our newly discovered loci with prostate-specific antigen (PSA) levels by using a series of disease-free controls ($n = 9{,}090$; Methods). Among the 48 available loci, we observed a significant association for rs8093601 ($P = 5.0 \times 10^{-4}$; Supplementary Table 13) after correction for multiple testing ($P = 0.05/48 = 1.0 \times 10^{-3}$). This marker lies near *MBD2* (encoding methyl-CpG binding domain protein 2) and has not

previously been associated with either PrCa risk or PSA levels. The effect estimates of PrCa clinical features and overall PrCa did not differ (Supplementary Table 14). LD plots incorporating several functional annotation features for each of the 63 novel markers are presented in Supplementary Fig. 4.

Several strong candidate genes were identified among the PrCa-susceptibility loci, including *ATM*, a key gene within the DNA-damage response pathway, in which truncating variants contribute to PrCa susceptibility and progression, particularly aggressive PrCa[34,35]. The index variant within this region is the missense variant rs1800057, exerting a modestly increased risk of PrCa (OR = 1.16; $P = 8.15 \times 10^{-9}$; G>C, p.Pro1054Arg; Fig. 2a). Although rs1800057 is designated 'benign' by ClinVar (see URLs), it has been suggested to be associated with a twofold-increased risk of early-onset PrCa in a small clinical series and has been found to be unassociated with morbidity after treatment[36]. In addition to the *ATM* region, we identified missense variants at three separate loci: rs2066827 within *CDKN1B,* encoding a cyclin-dependent-kinase inhibitor that controls cell-cycle progression; rs33984059 within *RFX7,* encoding a transcription factor; and rs2277283 within *INCENP*, encoding a centromere-interacting protein.

rs1048169 at 9p22 is located in the 3′ UTR of *HAUS6* (Fig. 2b), which encodes a subunit of augmin, a protein complex required for proper microtubule formation and chromosome segregation during cell division[37]. rs1048169 is also an eQTL for *HAUS6* expression. Interestingly, an additional lead SNP identified in this study, rs11666569 at 19p13, was found to be an eQTL for two genes, including *HAUS8,* which encodes another member of the augmin complex. These discoveries may implicate a potential role of augmin in PrCa susceptibility.

rs7968403 (OR = 1.06; $P = 3.38 \times 10^{-12}$; Fig. 2c) is situated within the first intron of *RASSF3.* Members of the Ras-association-domain family (RASSF) are putative tumor suppressors implicated in a range of biological processes[38]. RASSF3 is ubiquitously expressed across tissue types and has been observed to arrest the cell cycle in the G1 phase and to induce apoptosis through the p53 pathway[39]. A PrCa-risk locus, ~100 kb away, within *RASSF6* has been identified in a previous study[11]. However, rs7968403 was also an eQTL for the distant *WIF1* (encoding WNT-inhibitory factor 1; Fig. 2c). *WIF1* inhibits Wnt signaling and is frequently downregulated in PrCa[40], whereas aberrant activation of Wnt signaling is common in many solid tumor types. Restoration of *WIF1* expression has also been demonstrated to decrease cell motility and invasiveness in a metastatic PrCa cell line and to reduce tumor growth in a mouse xenograft model[41]. Both *RASSF3* and *WIF1* therefore are plausible mechanisms for the modulation of PrCa risk at this locus.

rs28441558 at 17p13 was the lead variant for a cluster of highly correlated SNPs centered on the *CHD3* gene (Fig. 2d). *CHD3* encodes an ATPase that forms a component of the nucleosome-remodeling and deacetylase (NuRD) histone deacetylase complex, which is involved in chromatin remodeling. NuRD plays an important role in regulating gene expression, as both a silencer and an activator of transcription, in addition to its roles in

URLs. ClinVar, http://www.ncbi.nlm.nih.gov/clinvar/.

maintaining genomic integrity and in the DNA-damage response[42]. Alterations in NuRD function have been implicated in several cancer types and found to act in a highly complex manner[43,44]. However, rs28441558 was also observed to be an eQTL for three genes: *LOC284023,* encoding a currently uncharacterized noncoding RNA transcript; *GUCY2D,* encoding a guanylate cyclase enzyme expressed predominantly in the retina; and *ALOX15B,* encoding a member of the lipoxygenase family of enzymes that produce fatty acid hyperoxides. Although *CHD3* appears to be the most biologically plausible candidate gene for this locus, we cannot exclude roles of any of these genes.

Our pathway analysis based on mapping each SNP to the nearest gene (Methods) by using the meta-analysis summary association statistic identified several pathways implicated in PrCa susceptibility. The top 53 pathways detected (enrichment score (ES) >0.50) are provided in Supplementary Table 15. The most significant pathway detected was PD-1 signaling (ID: 389948), ES = 0.74, as defined by the REACTOME database (Supplementary Fig. 5). This pathway is intriguing, given the therapeutic potential of several checkpoint inhibitors focusing on the PD-1 signaling pathway to enhance immune responses[45].

In summary, we identified 63 novel PrCa-susceptibility variants, including strong candidate loci highlighting the DNA-repair and cell-cycle pathways. Previous studies have probably overestimated the effect estimates of PrCa loci as a result of the 'winner's curse', thus yielding a biased familial relative risk (FRR) and polygenic risk score (PRS). Here, we applied a weighted Bayesian correction approach and demonstrated that our large sample size minimized the winner's curse bias[46] (Methods and Supplementary Fig. 6). We applied the beta estimates calculated in our overall meta-analysis to the OncoArray sample set to calculate the FRR and PRS risk models (Supplementary Table 16). Our prediction models included 85 previously reported PrCa loci replicating in our overall meta-analysis and our 62 novel loci associated with overall PrCa risk. Assuming a familial risk estimate of 2.5 for PrCa[47,48], we demonstrated that our 147 loci captured 28.4% of the FRR (Supplementary Table 17). The 62 newly identified PrCa loci increased the FRR by 4.4%. On the basis of the assumption of a log-additive model, the estimated RR for PrCa relative to men in the twenty-fifth to seventy-fifth PRS percentiles (baseline group) was 5.71 (95% CI: 5.04–6.48) for men in the top first percentile of the PRS distribution and 2.69 (95% CI: 2.55–2.82) for individuals in the ninetieth to ninety-ninth percentiles of the PRS distribution (Table 2). The PRS score was positively associated with overall PrCa risk (OR = 1.86; 95% CI: 1.83–1.89; Supplementary Table 18). Our novel associations highlight several biological pathways that warrant further investigation. The increased PRS can be used to improve the identification of men at high risk for PrCa and therefore inform PSA guidelines for screening and management to reduce the burden of over-testing.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0142-8.

Methods

Study subjects

A brief overview and study details for participating PrCa studies in the newly genotyped OncoArray project are provided in Supplementary Table 1 for men of European ancestry. All studies were approved by the appropriate ethics committees (as described in the references for each study listed in Supplementary Table 1), and informed consent was obtained from all participants. Supplementary Table 2 summarizes the PrCa sample series of the Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) consortium contributing both newly obtained genotyping data for the OncoArray and previous GWAS. Most of the studies contributing to the OncoArray were case-control studies primarily based in either the United States or Europe. In total, 52 new studies provided core data on disease status, age at diagnosis (age at observation or questionnaire for controls), family history of PrCa, and clinical factors for cases (for example, PSA at diagnosis and Gleason score) for 48,455 PrCa cases and 28,321 disease-free controls. Previous GWAS contributed an additional 32,255 PrCa cases and 33,202 disease-free controls of European ancestry to the overall meta-analysis[12]. Supplementary Table 3 provides QC information by consortia (e.g., OncoArray project, UK GWAS, and so forth) for both samples and SNPs. After removal of all overlapping samples, the OncoArray contribution for newly genotyped samples was 46,939 PrCa cases and 27,910 disease-free controls.

Several strata-specific analyses were implemented to evaluate the effects of genetic variation on PrCa disease aggressiveness. Supplementary Table 4 describes the analysis title, outcome and reference groups, and the statistical model used. Several classification schemes (low aggressiveness, intermediate aggressiveness, and so forth) were implemented to better assess the spectrum of genetic involvement. All classification schemes incorporated the diagnostic clinical features PSA, tumor stage, and Gleason score. To compare the results with those from previous PrCa aggressive analyses[12] by our research group, we included the ‘advanced (plus death due to PrCa)’ classification. Contributing study groups missing clinical features were excluded (Supplementary Table 2). Individuals with missing or granular clinical information were excluded. The strata-specific sample sizes from the PrCa GWAS consortium are provided in Supplementary Table 5. Furthermore, we analyzed Gleason score as a continuous variable.

OncoArray SNP selection

The NCI GAME-ON consortium (http://epi.grants.cancer.gov/gameon/) provided SNPs to be included in the Illumina OncoArray. Approximately 50% of the OncoArray was a compilation of SNP lists by the GAME-ON disease consortium of cancer (breast, colorectal, lung, ovarian, and prostate), a common set of variants for common risk regions, other related traits (BMI, age at menarche, and so forth), pharmacogenetics, and candidates[30]. The remaining content of the OncoArray was selected as a ‘GWAS backbone’ (Illumina HumanCore), which aimed to provide high coverage for most common variants through imputation. Approximately 79,000 SNPs were selected specifically for their relevance to PrCa, on the basis of prior evidence of association with overall or subtype-specific disease, fine-mapping of known PrCa regions, and candidate submissions (survival, exome

sequencing, and so forth). To maximize the efficiency of the array, cancer-specific candidate lists were merged to remove redundant genetic variation[30].

Genotype calling and quality control

Details of the genotype calling and QC for the iCOGS and GWAS have been described elsewhere[11–28].

Of the 568,712 variants selected for genotyping on the OncoArray, 533,631 were successfully manufactured on the array (including 778 duplicate probes). OncoArray genotyping of ELLIPSE studies was conducted at five sites (Cambridge, CIDR, Copenhagen, USC, and NCI). Details of the genotype calling for the OncoArray have been described in more detail elsewhere[30]. Briefly, we developed a single calling pipeline that was applied to more than 500,000 samples across the GAME-ON consortium. An initial cluster file was generated by using 56,284 samples selected from all major genotyping centers and ancestries, with the Gentrain2 algorithm. Variants likely to have problematic clusters were selected for manual inspection on the basis of the following criteria: call rate <99%, MAF <0.001, poor Illumina intensity and clustering metrics, deviation from the MAF observed in the 1KGP, by using the criterion $\frac{(|p_1 - p_0| - 0.01)^2}{((p_1 + p_0)(2 - p_1 - p_0))} > C$, where p_0 and p_1 are the minor frequencies in the 1KGP and OncoArray datasets, respectively, and $C = 0.008$. This procedure resulted in manual adjustment of the cluster file for 3,964 variants and the exclusion of 16,526 variants. The final cluster file was then applied to the full dataset.

Our QC pipeline for ELLIPSE excluded SNPs with a call rate <95% by study, not in Hardy-Weinberg equilibrium ($P < 10^{-7}$ in controls or $P < 10^{-12}$ in cases) or with concordance <98% among 11,260 duplicate pairs. To minimize imputation errors, we additionally excluded SNPs with a MAF <1% and a call rate <98% in any study, SNPs that could not be linked to the 1KGP reference, those with MAF for Europeans that differed from that for the 1KGP, and a further 16,526 SNPs for which the cluster plot was judged to be not ideal. Of the 533,631 manufactured SNPs on the OncoArray, we retained 498,417 SNPs among our samples of European ancestry after QC.

We excluded duplicate samples and first-degree relatives within each study, duplicates across studies, samples with a call rate <95%, and samples with extreme heterozygosity (>4.9 s.d. from the mean for the reported ancestry). We excluded duplicated samples as well as first-degree relatives across the GWAS studies CAPS1, CAPS2, UK Stage 1, UK Stage 2, and iCOGS. Duplicate and first-degree-related samples were assessed across the BPC3 and Pegasus GWAS studies as well. Ancestry was computed through principal component analysis using 2,318 informative markers on a subset of ~47,000 samples and projected onto the complete OncoArray dataset. The current analysis was restricted to men of European ancestry, defined as individuals with an estimated proportion of European ancestry >0.8, with reference to the HapMap populations, on the basis of the first two principal components. Of the 78,182 samples genotyped (regardless of ancestry), the final dataset consisted of 74,849 samples, of which 46,939 PrCa cases and 27,910 disease-free controls

(Supplementary Table 3), after exclusion of overlap samples, were meta-analyzed with previous studies.

Imputation

We imputed genotypes for ~70 million SNPs for all samples by using the October 2014 (Phase 3) release of the 1KGP data as the reference panel. We imputed the OncoArray and GWAS datasets through a two-stage imputation approach, using SHAPEIT[31] for phasing and IMPUTEv2 (ref. [32]) for imputation. The imputation was performed in 5-Mb nonoverlapping intervals. All subjects were split into subsets of ~10,000 samples, with subjects from the same group in the subset. We imputed genotypes for all SNPs that were polymorphic (MAF > 0.1%) in European samples. We excluded data for all monomorphic SNPs and those with an imputation r^2 <0.3, thus leaving a total of 20,370,935 SNPs across chromosomes 1–22 and chromosome X. Of the SNPs imputed, 49.3% had a MAF <1%, 15.2% had a MAF ranging between 1% and 5%, and 35.5% had a MAF ≥5%.

Statistical analyses

Per-allele odds ratios and standard errors were generated for the OncoArray and each GWAS, with adjustment for principal components and study-relevant covariates through logistic regression. The OncoArray and iCOGS analyses were additionally stratified by country and study, respectively. We used the first seven principal components in our analysis of individuals of European ancestry, because additional components did not further decrease inflation in the test statistics.

OR estimates were derived with either SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) or an in-house C++ program (Supplementary Table 3). OR estimates and standard errors were combined by a fixed-effects inverse variance meta-analysis in METAL[50]. All statistical tests conducted were two sided.

Our analyses included overall PrCa and several clinically relevant strata. These strata comprised: (i) high versus low aggressive PrCa; (ii) high versus low/intermediate aggressive PrCa; (iii) advanced versus nonadvanced PrCa; (iv) advanced PrCa versus controls; (v) early-onset PrCa (≤55 years) versus controls; and (iv) Gleason score (Supplementary Tables 4 and 5). We defined low aggressive as tumor stage ≤T1 and Gleason score ≤6 and PSA <10 ng/mL; intermediate aggressive as tumor stage T2 or Gleason score = 7 or PSA 10–20 ng/mL; high aggressive as tumor stage T3/T4 or N1 or M1, or Gleason score ≥8 or PSA >20 ng/mL; and advanced as either metastatic disease, Gleason score ≥8, PSA > 100 or PrCa-related death (Supplementary Table 4).

Definition of newly associated loci

To search for novel loci, we assessed all SNPs excluding those within a known PrCa locus, defined by current fine-mapping assessments (Supplementary Table 7). SNPs that were associated with disease risk at $P < 5 \times 10^{-8}$ in the meta-analysis (GWAS and OncoArray) were considered novel. The SNP with the lowest P value in a region was considered the lead SNP. Imputation quality was assessed on the basis of IMPUTE2 imputation r^2 in the OncoArray dataset (Supplementary Table 8).

For ten regions where the newly identified locus was near a previously known region, we reported a novel association if the pairwise r^2 between the new and the previously known SNP was <0.2. For novel PrCa associations for which the variant was imputed in the OncoArray study sample series and had an imputed quality score <0.70, we assessed the quality of the imputation by masking the variant in a subset of the 1KGP European sample and calculating the concordance after reimputation in the remaining 1KGP samples.

Reliability of imputation

Novel SNPs with an IMPUTE2 r^2 <0.80 among the OncoArray sample series (Supplementary Table 8) were flagged for further investigation to minimize the probability of false positives. First, we examined LD plots (http://locuszoom.org/) for poorly imputed SNPs (±500 kb), including only genotyped SNPs within the region. The imputed index SNP was included in the plot to determine the strength of LD with nearby signals and to assess a pattern of association. Furthermore, we performed an imputation experiment using the 2,504 1KGP Phase 3 samples. We split this sample into two parts: a random sample of 259 individuals of European ancestry (excluding Finnish individuals) and a mixed-population reference panel of 2,245 individuals. The random sample of 259 individuals of European ancestry was filtered to include only the genetic variants available from the OncoArray after QC. This procedure ensured that the same imputation input was used in the overall imputation. The 259 individuals were imputed by using 2,245 individuals as the reference panel. A 5-Mb segment of the genome was selected on the basis of the target SNP (±250 Mb). SHAPEIT2 was used for prephasing, and IMPUTE2 was used for imputation. Customized imputation settings included an effect size of 20,000, allowance of large-region imputation and a random seed of 12345. A weighted linear kappa statistic was calculated to determine the correlation of the imputation with the true genotypes.

We evaluated four SNPs whose IMPUTE2 r^2 was <0.80 in the OncoArray sample series: rs527510716 (chr 7), rs6602880 (chr 10), rs533722308 (chr 18), and rs144166867 (chr X). Supplementary Fig. 3 includes the LD plots for three of the poorly imputed SNPs. The variant rs144166867 (chr X) could not be plotted, because no genotype SNPs were available within ±500 kb on the OncoArray. Both LD plots for markers rs527510716 (chr 7) and rs533722308 (chr 18) showed significant associations ($P < 1 \times 10^{-3}$) for several genotype markers with moderate LD of the index SNP The kappa coefficients for markers rs527510716 (chr 7) and rs533722308 (chr 18) were 0.911 and 0.931, respectively (Supplementary Table 9). The marker rs6602880 (chr 10) had a kappa coefficient of 0.812 and was the only significant variant in the LD plot. The kappa coefficient for marker rs144166867 (chr X) was 0.665 (Supplementary Table 9). The markers rs6602880 (chr 10) and rs144166867 (chr X) were probably false positives due to poor imputation for these regions.

Proportion of familial risk explained

The contribution of the known SNPs to the familial risk of PrCa, under a multiplicative model, was computed with the formula

$$\sum_k (\log \lambda_k)/(\log \lambda_0)$$

where λ_0 is the observed familial risk to first-degree relatives of PrCa cases[47,48], assumed to be 2.5, and λ_k is the familial relative risk due to locus *k,* given by:

$$\lambda_k = \frac{p_k r_k^2 + q_k}{(p_k r_k + q_k)^2}$$

where p_k is the frequency of the risk allele for locus *k,* $q_k = 1 - p_k$, and r_k is the estimated per-allele odds ratio.

On the basis of the assumption of a log-additive model, we constructed a PRS from the summed risk-allelic doses weighted by the per-allele log ORs. Thus, for each individual *j,* we derived:

$$score_j = \sum_{i=1}^{N} \beta_i g_{ij}$$

where N is the number of SNPs, g_{ij} is the allele dose at SNP_i for individual *j,* and β_i is the per-allele log-odds ratio of SNP_i.

The risk of PrCa was estimated for the percentiles of the distribution of the PRS (<1, 1–10, 10–25, 25–75, 75–90, 90–99, >99 and <10, 10–25, 25–75, 75–90, >90) for which cumulative score thresholds were determined according to the observed distribution among controls. We applied effect sizes and allele frequencies obtained from the overall meta-analysis of Europeans to estimate risk scores for individuals of European ancestry in the OncoArray study[51]. A standardized PRS score was calculated by dividing the observed PRS score by the s.d. of the PRS score among controls. A logistic-regression framework was used to evaluate the percentile comparisons and to determine the risk estimate. The models were adjusted for the first seven principal components to account for population stratification and stratified by country.

The FRR and PRS risk estimation was limited to the variants for which our overall meta-analysis indicated a statistically significant association. In total, we included 147 PrCa index SNPs in our risk-score modeling, including 85 previously published associations and the 62 novel findings reported here. To correct for potential bias in effect estimation of newly discovered variants, we implemented a fully Bayesian version of a weighted correction given in equation (3).4 in ref. [46]. Specifically, we placed a normal prior distribution on MLE effect estimates of the form $\beta_m \sim N(\beta_{Cor}, \tau^2)$. Here, β_m is the log OR from the overall meta-analysis; β_{Cor} is the bias-corrected estimate calculated with the expectation-adjusted estimator from equation (3).1 in in ref. [46]; and τ is a prespecified variance of the effect distribution reflecting the bias and is defined as $\tau = \left|\hat{\beta}_m - \beta_{Cor}\right|$.

eQTL analyses

Genotype and gene expression data were downloaded from TCGA for 494 samples with PrCa (https://gdc-portal.nci.nih.gov/). QC was performed on both these datasets as follows: on the genotype, we filtered out samples with high heterozygosity (mean heterozygosity ±2 s.d.) and missing genotypes and duplicated or related samples. We then performed principal component analysis on the 494 samples plus 2,506 samples from 1KGP to infer the ancestry of the TCGA samples; samples of non-European ancestry were removed. We also filtered out variants with missing call rate >5%. For the expression data, samples from two plates had, on average, much higher expression values than did the remaining samples and therefore were excluded. We also filtered genes with mean expression across samples ≤6 counts. Finally, expression values were quantile-normalized by samples and rank-transformed by genes. After QC, we used the data from 359 samples. For the eQTL analysis, 35 PEER factors from the top 10,000 expressed genes were used as covariates, plus three genotyping PCs (which explained 18% of total variation). eQTL analysis was performed in FastQTL with 1,000 permutations over the 85 regions. We used a window of 1 Mb (upstream/downstream) from the transcription start site of each gene.

Gene set enrichment analyses

The file Human_GOBP_AllPathways_no_GO_iea_September_01_2016_symbol.gmt (http://baderlab.org/GeneSets/) from the GeneSets database[52], was used for all analyses. This database contains pathways from Reactome[53], NCI Pathway Interaction Database[54], Gene Ontology (GO) biological process[55], HumanCyc[56], MSigdb[57], NetPath[58], and Panther[59]. We manually corrected several pathways in which the *c* gene was entered as *PDK1*. GO pathways inferred from electronic annotation terms were excluded. The same pathway (for example, apoptosis) may be defined in two or more databases with potentially different sets of genes, and all versions of these duplicate/overlapping pathways were included. Pathway size was determined by the total number of genes in the pathway to which SNPs in the imputed GWAS dataset could be mapped. To provide more biologically meaningful results, and to reduce false positives, only pathways that contained between 10 and 200 genes were considered.

Gene information (hg19) was downloaded from the ANNOVAR[60] website (http://annovar.openbioinformatics.org/). SNPs were mapped to the nearest gene within 500-kb windows; those that were further away from any gene were excluded. Gene significance was calculated by assigning the lowest *P* value observed across all SNPs assigned to a gene[61,62], on the basis of the combined European meta-analysis (previous GWAS and OncoArray).

The gene-set enrichment analysis (GSEA)[52] algorithm, as implemented in the GenGen package (http://gengen.openbioinformatics.org/en/latest/)[62,63], was used to perform pathway analysis. Briefly, the algorithm calculates an ES for each pathway on the basis of a weighted Kolmogorov-Smirnov statistic[63]. To calculate the ES, we performed 100 permutations and averaged the final score. Pathways with most of their genes at the top of the ranked list of genes obtain higher ES values. Only pathways with positive ES and at least one gene with $P < 5 \times 10^{-8}$ were retained for subsequent analysis. An enrichment map was created in the Enrichment Map (EM) v 2.1.0 application[52] in Cytoscape v3.40 (ref. [64]), with application of

force-directed layout, in weighted mode. We restricted our pathway analysis to those with an ES ≥0.50 to ensure a true-positive rate >0.20 and a false-positive rate <0.15.

Reporting Summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The OncoArray genotype data and relevant covariate information (ancestry, country, principal components, and so forth) generated during this study have been deposited in dbGaP under accession code phs001391.v1.p1. In total, 47 of the 52 OncoArray studies encompassing nearly 90% of the individual samples will be available (Supplementary Table 19). The previous meta-analysis summary results and genotype data[12] are available in dbGaP under accession code phs001081.v1.p1. The complete meta-analysis summary associations statistics are publicly available at the PRACTICAL website (http://practical.icr.ac.uk/blog/).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Fredrick R. Schumacher[1,2,139,140,*], Ali Amin Al Olama[3,4,139,140,*], Sonja I. Berndt[5,139,140], Sara Benlloch[3,6], Mahbubl Ahmed[6], Edward J. Saunders[6], Tokhir Dadaev[6], Daniel Leongamornlert[6], Ezequiel Anokian[6], Clara Cieza-Borrella[6], Chee Goh[6], Mark N. Brook[6], Xin Sheng[7], Laura Fachal[8,9], Joe Dennis[3], Jonathan Tyrer[3], Kenneth Muir[10,11], Artitaya Lophatananon[10,11], Victoria L. Stevens[12], Susan M. Gapstur[12], Brian D. Carter[12], Catherine M. Tangen[13], Phyllis J. Goodman[13], Ian M. Thompson Jr[14], Jyotsna Batra[15,16], Suzanne Chambers[17,18], Leire Moya[15,16], Judith Clements[15,16], Lisa Horvath[19,20], Wayne Tilley[21], Gail P. Risbridger[22,23], Henrik Gronberg[24], Markus Aly[25,26], Tobias Nordström[24,27], Paul Pharoah[3,9], Nora Pashayan[9,28], Johanna Schleutker[29,30], Teuvo L. J. Tammela[31], Csilla Sipeky[29], Anssi Auvinen[32], Demetrius Albanes[5], Stephanie Weinstein[5], Alicja Wolk[33,34], Niclas Håkansson[33], Catharine M. L. West[35], Alison M. Dunning[9], Neil Burnet[36], Lorelei A. Mucci[37], Edward Giovannucci[37], Gerald L. Andriole[38], Olivier Cussenot[39,40], Géraldine Cancel-Tassin[39,40], Stella Koutros[5], Laura E. Beane Freeman[5], Karina Dalsgaard Sorensen[41,42], Torben Falck Orntoft[41,42], Michael Borre[42,43], Lovise Maehle[44], Eli Marie Grindedal[44], David E. Neal[45,46,47], Jenny L. Donovan[48], Freddie C. Hamdy[47], Richard M. Martin[48], Ruth C. Travis[49], Tim J. Key[49], Robert J. Hamilton[50], Neil E. Fleshner[50], Antonio Finelli[51], Sue Ann Ingles[7], Mariana C. Stern[7], Barry S. Rosenstein[52,53], Sarah L. Kerns[54], Harry Ostrer[55], Yong-Jie Lu[56], Hong-Wei Zhang[57], Ninghan Feng[58], Xueying Mao[56], Xin Guo[59], Guomin Wang[60], Zan Sun[61], Graham G. Giles[62,63], Melissa C. Southey[64], Robert J. MacInnis[62,63], Liesel M. FitzGerald[62,65], Adam S. Kibel[66], Bettina F. Drake[38], Ana

Vega[67], Antonio Gómez-Caamaño[68], Robert Szulkin[69,70], Martin Eklund[24], Manolis Kogevinas[71,72,73,74], Javier Llorca[72,75], Gemma Castaño-Vinyals[71,72,73,74], Kathryn L. Penney[76], Meir Stampfer[76], Jong Y. Park[77], Thomas A. Sellers[77], Hui-Yi Lin[78], Janet L. Stanford[79,80], Cezary Cybulski[81], Dominika Wokolorczyk[81], Jan Lubinski[81], Elaine A. Ostrander[82], Milan S. Geybels[79], Børge G. Nordestgaard[83,84], Sune F. Nielsen[83,84], Maren Weischer[84], Rasmus Bisbjerg[85], Martin Andreas Røder[86], Peter Iversen[86], Hermann Brenner[87,88,89], Katarina Cuk[87], Bernd Holleczek[90], Christiane Maier[91], Manuel Luedeke[91], Thomas Schnoeller[92], Jeri Kim[93], Christopher J. Logothetis[93], Esther M. John[94,95], Manuel R. Teixeira[96,97], Paula Paulo[96], Marta Cardoso[96], Susan L. Neuhausen[98], Linda Steele[98], Yuan Chun Ding[98], Kim De Ruyck[99], Gert De Meerleer[99], Piet Ost[100], Azad Razack[101], Jasmine Lim[101], Soo-Hwang Teo[102], Daniel W. Lin[79,103], Lisa F. Newcomb[79,103], Davor Lessel[104], Marija Gamulin[105], Tomislav Kulis[106], Radka Kaneva[107], Nawaid Usmani[108,109], Sandeep Singhal[108,109], Chavdar Slavov[110], Vanio Mitev[107], Matthew Parliament[108,109], Frank Claessens[111], Steven Joniau[112], Thomas Van den Broeck[111,112], Samantha Larkin[113], Paul A. Townsend[114], Claire Aukim-Hastie[115], Manuela Gago-Dominguez[116,117], Jose Esteban Castelao[118], Maria Elena Martinez[119], Monique J. Roobol[120], Guido Jenster[120], Ron H. N. van Schaik[121], Florence Menegaux[122], Thérèse Truong[122], Yves Akoli Koudou[122], The Profile Study[123], Jianfeng Xu[124], Kay-Tee Khaw[125], Lisa Cannon-Albright[126,127], Hardev Pandha[115], Agnieszka Michael[115], Stephen N. Thibodeau[128], Shannon K. McDonnell[129], Daniel J. Schaid[129], Sara Lindstrom[130], Constance Turman[131], Jing Ma[76], David J. Hunter[131], Elio Riboli[132], Afshan Siddiq[133], Federico Canzian[134], Laurence N. Kolonel[135], Loic Le Marchand[135], Robert N. Hoover[5], Mitchell J. Machiela[5], Zuxi Cui[1], Peter Kraft[131], Australian Prostate Cancer BioResource (APCB)[123], The IMPACT Study[123], Canary PASS Investigators[123], Breast and Prostate Cancer Cohort Consortium (BPC3)[123], The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium[123], Cancer of the Prostate in Sweden (CAPS)[123], Prostate Cancer Genome-wide Association Study of Uncommon Susceptibility Loci (PEGASUS)[123], The Genetic Associations and Mechanisms in Oncology (GAME-ON)/Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) Consortium[123], Christopher I. Amos[136,137,138], David V. Conti[7,141], Douglas F. Easton[3,9,141], Fredrik Wiklund[24,141], Stephen J. Chanock[5,141], Brian E. Henderson[7,141,142], Zsofia Kote-Jarai[6,141], Christopher A. Haiman[7,141], and Rosalind A. Eeles[6,139,141,*]

Affiliations

[1]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA. [2]Seidman Cancer Center, University Hospitals, Cleveland, OH, USA. [3]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK. [4]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA. [6]Institute of Cancer Research, London, UK. [7]Department of Preventive Medicine, Keck School of

Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA. [8]Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica, CIBERER, IDIS, Santiago de Compostela, Spain. [9]Department of Oncology, Centre for Cancer Genetic Epidemiology, Strangeways Laboratory, University of Cambridge, Cambridge, UK. [10]Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK. [11]Warwick Medical School, University of Warwick, Coventry, UK. [12]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA. [13]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [14]CHRISTUS Santa Rosa Hospital-Medical Center, San Antonio, TX, USA. [15]Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia. [16]Australian Prostate Cancer Research Centre-Qld, Translational Research Institute, Brisbane, Queensland, Australia. [17]Menzies Health Institute of Queensland, Griffith University, Nathan, Queensland, Australia. [18]Cancer Council Queensland, Fortitude Valley, Queensland, Australia. [19]Chris O'Brien Lifehouse (COBLH), Camperdown, New South Wales, Australia. [20]Garvan Institute of Medical Research, Sydney, New South Wales, Australia. [21]Dame Roma Mitchell Cancer Research Centre, University of Adelaide, Adelaide, South Australia, Australia. [22]Monash Biomedicine Discovery Institute Cancer Program, Prostate Cancer Research Program, Department of Anatomy and Developmental Biology, Monash University, Victoria, Australia. [23]Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia. [24]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [25]Department of Urology, Karolinska University Hospital, Stockholm, Sweden. [26]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [27]Department of Clinical Sciences at Danderyds Hospital, Karolinska Institutet, Stockholm, Sweden. [28]Department of Applied Health Research, University College London, London, UK. [29]Institute of Biomedicine, University of Turku, Turku, Finland. [30]Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Turku, Finland. [31]Department of Urology, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [32]Department of Epidemiology, School of Health Sciences, University of Tampere, Tampere, Finland. [33]Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [34]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden. [35]Division of Cancer Sciences, Manchester Cancer Research Centre, Manchester Academic Health Science Centre and Christie NHS Foundation Trust, University of Manchester, Manchester, UK. [36]Department of Oncology, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, University of Cambridge, Cambridge, UK. [37]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [38]Washington University School of Medicine, St. Louis, MO, USA. [39]UPMC Sorbonne Universités, GRC no. 5 ONCOTYPE-URO, Tenon Hospital, Paris, France. [40]CeRePP, Tenon Hospital, Paris, France. [41]Department of Molecular Medicine,

Aarhus University Hospital, Aarhus, Denmark. [42]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. [43]Department of Urology, Aarhus University Hospital, Aarhus, Denmark. [44]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [45]Department of Oncology, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK. [46]Cancer Research UK, Cambridge Research Institute, Cambridge, UK. [47]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, Faculty of Medical Science, John Radcliffe Hospital, University of Oxford, Oxford, UK. [48]School of Social and Community Medicine, University of Bristol, Bristol, UK. [49]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. [50]Department of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [51]Division of Urology, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [52]Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [53]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [54]Department of Radiation Oncology, University of Rochester Medical Center, Rochester, NY, USA. [55]Department of Pathology, Albert Einstein College of Medicine, Bronx, NY, USA. [56]Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK. [57]Second Military Medical University, Shanghai, China. [58]Wuxi Second Hospital, Nanjing Medical University, Wuxi, China. [59]Department of Urology, First Affiliated Hospital, Chongqing Medical University, Chongqing, China. [60]Department of Urology, Zhongshan Hospital, Fudan University Medical College, Shanghai, China. [61]People's Hospital of Liaoning Province, People's Hospital of China Medical University, Shenyang, China. [62]Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia. [63]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia. [64]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia. [65]Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia. [66]Division of Urologic Surgery, Brigham and Womens Hospital, Boston, MA, USA. [67]Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica, CIBERER, IDIS, Santiago de Compostela, Spain. [68]Department of Radiation Oncology, Complexo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain. [69]Division of Family Medicine, Department of Neurobiology, Care Science and Society, Karolinska, Institutet, Huddinge, Sweden. [70]Scandinavian Development Services, Danderyd, Sweden. [71]ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. [72]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [73]IMIM (Hospital del Mar Research Institute), Barcelona, Spain. [74]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [75]University of Cantabria-IDIVAL, Santander, Spain. [76]Department of Medicine, Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [77]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA. [78]Biostatistics Program, School of Public Health, Louisiana State University Health Sciences

Center, New Orleans, LA, USA. [79]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [80]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA. [81]International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. [82]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [83]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [84]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [85]Department of Urology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [86]Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. [87]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [88]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [89]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. [90]Saarland Cancer Registry, Saarbrücken, Germany. [91]Institute for Human Genetics, University Hospital Ulm, Ulm, Germany. [92]Department of Urology, University Hospital Ulm, Ulm, Germany. [93]Department of Genitourinary Medical Oncology, University of Texas-MD Anderson Cancer Center, Houston, TX, USA. [94]Cancer Prevention Institute of California, Fremont, CA, USA. [95]Department of Health Research & Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. [96]Department of Genetics, Portuguese Oncology Institute of Porto, Porto, Portugal. [97]Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal. [98]Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA, USA. [99]Faculty of Medicine and Health Sciences, Basic Medical Sciences, Ghent University, Ghent, Belgium. [100]Department of Radiotherapy, Ghent University Hospital, Ghent, Belgium. [101]Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [102]Cancer Research Malaysia (CRM), Outpatient Centre, Subang Jaya Medical Centre, Selangor, Malaysia. [103]Department of Urology, University of Washington, Seattle, WA, USA. [104]Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [105]Department of Oncology, Division of Medical Oncology, Urogenital Unit, University Hospital Centre Zagreb, Zagreb, Croatia. [106]Department of Urology, University Hospital Center Zagreb, University of Zagreb School of Medicine, Zagreb, Croatia. [107]Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, Bulgaria. [108]Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Alberta, Canada. [109]Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta, Canada. [110]Department of Urology and Alexandrovska University Hospital, Medical University of Sofia, Sofia, Bulgaria. [111]Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium. [112]Department of Urology, University Hospitals Leuven, Leuven, Belgium.

[113]University of Southampton, Southampton General Hospital, Southampton, UK. [114]Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Center, NIHR Manchester Biomedical Research Centre, Health Innovation Manchester, University of Manchester, Manchester, UK. [115]University of Surrey, Guildford, Surrey, UK. [116]Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, Santiago de Compostela, Spain. [117]Moores Cancer Center, University of California San Diego, La Jolla, CA, USA. [118]Genetic Oncology Unit, CHUVI Hospital, Instituto de Investigación Biomédica Galicia Sur (IISGS), Complexo Hospitalario Universitario de Vigo, Vigo, Spain. [119]Department of Family Medicine and Public Health, Moores Cancer Center, University of California San Diego, La Jolla, CA, USA. [120]Department of Urology, Erasmus University Medical Center, Rotterdam, The Netherlands. [121]Department of Clinical Chemistry, Erasmus University Medical Center, Rotterdam, The Netherlands. [122]Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France. [123]A list of members and affiliations appears in the Supplementary Note. [124]Program for Personalized Cancer Care, NorthShore University HealthSystem, Evanston, IL, USA. [125]Clinical Gerontology Unit, University of Cambridge, Cambridge, UK. [126]Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA. [127]George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT, USA. [128]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. [129]Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA. [130]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA. [131]Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [132]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [133]Genomics England, Queen Mary University of London, London, UK. [134]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. [135]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA. [136]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA. [137]Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [138]Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA. [139]Royal Marsden NHS Foundation Trust, London, UK. [140]These authors contributed equally: Fredrick R. Schumacher, Ali Amin Al Olama, Sonja I. Berndt. [141]These authors jointly supervised this work: David V. Conti, Douglas F. Easton, Fredrik Wiklund, Stephen J. Chanock, Brian E. Henderson, Z. Sofia Kote-Jarai, Christopher A. Haiman, Rosalind A. Eeles. [142]Deceased: Brian E. Henderson.

Acknowledgements

We pay tribute to Brian Henderson for his vision and leadership; he was a driving force behind the OncoArray project and he unfortunately passed away before seeing its fruition. We also thank the individuals who participated in these studies enabling this work.

Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) (U19 CA 148537 for the ELLIPSE project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract no. HHSN268201200008I. Additional analytical support was provided by NIH NCI U01 CA188392 (to F.R.S.).

Funding for the iCOGS infrastructure came from the European Community's Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, and C8197/A16565), the NIH (CA128978) and Post-Cancer GWAS Initiative (1U19 CA148537, 1U19 CA148065, and 1U19 CA148112; the GAME-ON initiative), the Department of Defense (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

This work was supported by the Canadian Institutes of Health Research; the European Commission's Seventh Framework Programme grant agreement no. 223175 (HEALTH-F2-2009-223175); Cancer Research UK grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, and C16913/A6135; and NIH Cancer Post-Cancer GWAS initiative grant no. 1 U19 CA 148537-01 (the GAME-ON initiative).

We also thank the following for funding support: the Institute of Cancer Research and the Everyman Campaign, the Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), the Orchid Cancer Appeal, the National Cancer Research Network UK, and the National Cancer Research Institute (NCRI) UK. We are grateful for the support of NIHR funding to the NIHR Biomedical Research Centre at the Institute of Cancer Research and the Royal Marsden NHS Foundation Trust.

The Prostate Cancer Program of Cancer Council Victoria also acknowledges grant support from the National Health and Medical Research Council, Australia (126402, 209057, 251533, 396414, 450104, 504700, 504702, 504715, 623204, 940394, and 614296), VicHealth, Cancer Council Victoria, the Prostate Cancer Foundation of Australia, the Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. E.A.O., D.M.K., and E.M.K. acknowledge the Intramural Program of the National Human Genome Research Institute for support.

The BPC3 was supported by the NIH, National Cancer Institute (cooperative agreements U01-CA98233 to D.J.H., U01-CA98710 to S.M.G., U01-CA98216 to E.R., and U01-CA98758 to B.E.H., and the Intramural Research Program of the NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics).

The CAPS GWAS study was supported by the Swedish Cancer Foundation (grant nos. 09-0677, 11-484, and 12-823), the Cancer Risk Prediction Center (CRisP; http://ki.se/en/meb/crisp/), a Linneus Centre grant (contract ID 70867902) financed by the Swedish Research Council, and the Swedish Research Council (grant nos. K2010-70X-20430-04-3, and 2014-2269).

PEGASUS was supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH.

A full description of funding and acknowledgements can be found in the Supplementary Note.

References

1. Goh CL et al. Genetic variants associated with predisposition to prostate cancer and potential clinical implications. J. Intern. Med 271, 353–365 (2012). [PubMed: 22308973]
2. Siegel RL, Miller KD & Jemal A Cancer statistics, 2016. CA Cancer J. Clin 66, 7–30 (2016). [PubMed: 26742998]
3. Cuzick J et al. Prevention and early detection of prostate cancer. Lancet Oncol. 15, e484–e492 (2014). [PubMed: 25281467]
4. Altekruse SF et al. Spatial patterns of localized-stage prostate cancer incidence among white and black men in the southeastern United States, 1999–2001. Cancer Epidemiol. Biomarkers Prev 19, 1460–1467 (2010). [PubMed: 20501756]
5. Stanford JL & Ostrander EA Familial prostate cancer. Epidemiol. Rev 23, 19–23 (2001). [PubMed: 11588848]

6. Bunker CH et al. High prevalence of screening-detected prostate cancer among Afro-Caribbeans: the Tobago Prostate Cancer Survey. Cancer Epidemiol. Biomarkers Prev 11, 726–729 (2002). [PubMed: 12163325]
7. Ghadirian P, Howe GR, Hislop TG & Maisonneuve P Family history of prostate cancer: a multi-center case-control study in Canada. Int. J. Cancer 70, 679–681 (1997). [PubMed: 9096649]
8. Gronberg H, Damber L & Damber JE Familial prostate cancer in Sweden: a nationwide register cohort study. Cancer 77, 138–143 (1996). [PubMed: 8630920]
9. Matikaine MP et al. Relatives of prostate cancer patients have an increased risk of prostate and stomach cancers: a population-based, cancer registry study in Finland. Cancer Causes Control 12, 223–230 (2001). [PubMed: 11405327]
10. Eeles R et al. The genetic epidemiology of prostate cancer and its clinical implications. Nat. Rev. Urol 11, 18–31 (2014). [PubMed: 24296704]
11. Eeles RA et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat. Genet 45, 385–391 (2013). e1-e2. [PubMed: 23535732]
12. Al Olama AA et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat. Genet 46, 1103–1109 (2014). [PubMed: 25217961]
13. Al Olama AA et al. Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat. Genet 41, 1058–1060 (2009). [PubMed: 19767752]
14. Amundadottir LT et al. A common variant associated with prostate cancer in European and African populations. Nat. Genet 38, 652–658 (2006). [PubMed: 16682969]
15. Eeles RA et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat. Genet 41, 1116–1121 (2009). [PubMed: 19767753]
16. Eeles RA et al. Multiple newly identified loci associated with prostate cancer susceptibility. Nat. Genet. 40, 316–321 (2008). [PubMed: 18264097]
17. Gudmundsson J et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat. Genet 41, 1122–1126 (2009). [PubMed: 19767754]
18. Gudmundsson J et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat. Genet 39, 631–637 (2007). [PubMed: 17401366]
19. Gudmundsson J et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat. Genet. 40, 281–283 (2008). [PubMed: 18264098]
20. Gudmundsson J et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat. Genet 39, 977–983 (2007). [PubMed: 17603485]
21. Haiman CA et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. Nat. Genet. 43, 570–573 (2011). [PubMed: 21602798]
22. Kote-Jarai Z et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nat. Genet 43, 785–791 (2011). [PubMed: 21743467]
23. Schumacher FR et al. Genome-wide association study identifies new prostate cancer susceptibility loci. Hum. Mol. Genet 20, 3867–3875 (2011). [PubMed: 21743057]
24. Sun J et al. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat. Genet 40, 1153–1155 (2008). [PubMed: 18758462]
25. Takata R et al. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. Nat. Genet 42, 751–754 (2010). [PubMed: 20676098]
26. Thomas G et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat. Genet 40, 310–315 (2008). [PubMed: 18264096]
27. Yeager M et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat. Genet 39, 645–649 (2007). [PubMed: 17401363]
28. Duggan D et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J. Natl. Cancer Inst 99, 1836–1844 (2007). [PubMed: 18073375]
29. Al Olama Amin A. et al. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. Hum. Mol. Genet 22, 408–415 (2013). [PubMed: 23065704]

30. Amos CI et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. Cancer Epidemiol. Biomarkers Prev. 26, 126–135 (2017). [PubMed: 27697780]
31. Delaneau O, Marchini J & Zagury JF A linear complexity phasing method for thousands of genomes. Nat. Methods 9, 179–181 (2011). [PubMed: 22138821]
32. Howie BN, Donnelly P & Marchini J A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009). [PubMed: 19543373]
33. de Bakker PI et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128 (2008). [PubMed: 18852200]
34. Leongamornlert D et al. Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease. Br. J. Cancer 110, 1663–1672 (2014). [PubMed: 24556621]
35. Mateo J et al. DNA-repair defects and Olaparib in metastatic prostate cancer. N. Engl. J. Med 373, 1697–1708 (2015). [PubMed: 26510020]
36. Meyer A et al. ATM missense variant P1054R predisposes to prostate cancer. Radiother. Oncol 83, 283–288 (2007). [PubMed: 17502119]
37. Sanchez-Huertas C & Luders J The augmin connection in the geometry of microtubule networks. Curr. Biol 25, R294–R299 (2015). [PubMed: 25829017]
38. Volodko N, Gordon M, Salla M, Ghazaleh HA & Baksh S RASSF tumor suppressor gene family: biological functions and regulation. FEBS Lett. 588, 2671–2684 (2014). [PubMed: 24607545]
39. Kudo T et al. The RASSF3 candidate tumor suppressor induces apoptosis and G1-S cell-cycle arrest via p53. Cancer Res. 72, 2901–2911 (2012). [PubMed: 22593196]
40. Wissmann C et al. WIF1, a component of the Wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer. J. Pathol. 201, 204–212 (2003). [PubMed: 14517837]
41. Yee DS et al. The Wnt inhibitory factor 1 restoration in prostate cancer cells was associated with reduced tumor growth, decreased capacity of cell migration and invasion and a reversal of epithelial to mesenchymal transition. Mol. Cancer 9, 162 (2010). [PubMed: 20573255]
42. Allen HF, Wade PA & Kutateladze TG The NuRD architecture. Cell. Mol. Life Sci 70, 3513–3524 (2013). [PubMed: 23340908]
43. Lai AY & Wade PA Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat. Rev. Cancer 11, 588–596 (2011). [PubMed: 21734722]
44. Basta J & Rauchman M The nucleosome remodeling and deacetylase complex in development and disease. Transl. Res 165, 36–47 (2015). [PubMed: 24880148]
45. McDermott DF & Atkins MB PD-1 as a potential target in cancer therapy. Cancer Med. 2, 662–673 (2013). [PubMed: 24403232]
46. Zhong H & Prentice RL Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. Biostatistics 9, 621–634 (2008). [PubMed: 18310059]
47. Kiciński M, Vangronsveld J & Nawrot TS An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. PLoS One 6, e27130 (2011). [PubMed: 22073129]
48. Albright F et al. Prostate cancer risk prediction based on complete prostate cancer family history. Prostate 75, 390–398 (2015). [PubMed: 25408531]
49. Wang M et al. Large-scale association analysis in Asians identifies new susceptibility loci for prostate cancer. Nat. Commun 6, 8469 (2015). [PubMed: 26443449]
50. Willer CJ, Li Y & Abecasis GR METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 26, 2190–2191 (2010). [PubMed: 20616382]
51. Al Olama Amin A. et al. Risk analysis of prostate cancer in PRACTICAL, a multinational consortium, using 25 known prostate cancer susceptibility loci. Cancer Epidemiol. Biomarkers Prev 24, 1121–1129 (2015). [PubMed: 25837820]
52. Merico D, Isserlin R, Stueker O, Emili A & Bader GD Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5, e13984 (2010). [PubMed: 21085593]

53. Joshi-Tope G et al. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res. 33, D428–D432 (2005). [PubMed: 15608231]
54. Schaefer CF et al. PID: the Pathway Interaction Database. Nucleic Acids Res. 37, D674–D679 (2009). [PubMed: 18832364]
55. Ashburner M et al. Gene ontology: tool for the unification of biology. Nat. Genet 25, 25–29 (2000). [PubMed: 10802651]
56. Romero P et al. Computational prediction of human metabolic pathways from the complete human genome. Genome Biol. 6, R2 (2005). [PubMed: 15642094]
57. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550 (2005). [PubMed: 16199517]
58. Kandasamy K et al. NetPath: a public resource of curated signal transduction pathways. Genome Biol. 11, R3 (2010). [PubMed: 20067622]
59. Thomas PD et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13, 2129–2141 (2003). [PubMed: 12952881]
60. Wang K, Li M & Hakonarson H ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010). [PubMed: 20601685]
61. Wang L, Jia P, Wolfinger RD, Chen X & Zhao Z Gene set analysis of genome-wide association studies: methodological issues and perspectives. Genomics 98, 1–8 (2011). [PubMed: 21565265]
62. Wang K, Li M & Hakonarson H Analysing biological pathways in genome-wide association studies. Nat. Rev. Genet 11, 843–854 (2010). [PubMed: 21085203]
63. Wang K, Li M & Bucan M Pathway-based approaches for analysis of genomewide association studies. Am. J. Hum. Genet 81, 1278–1283 (2007). [PubMed: 17966091]
64. Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003). [PubMed: 14597658]

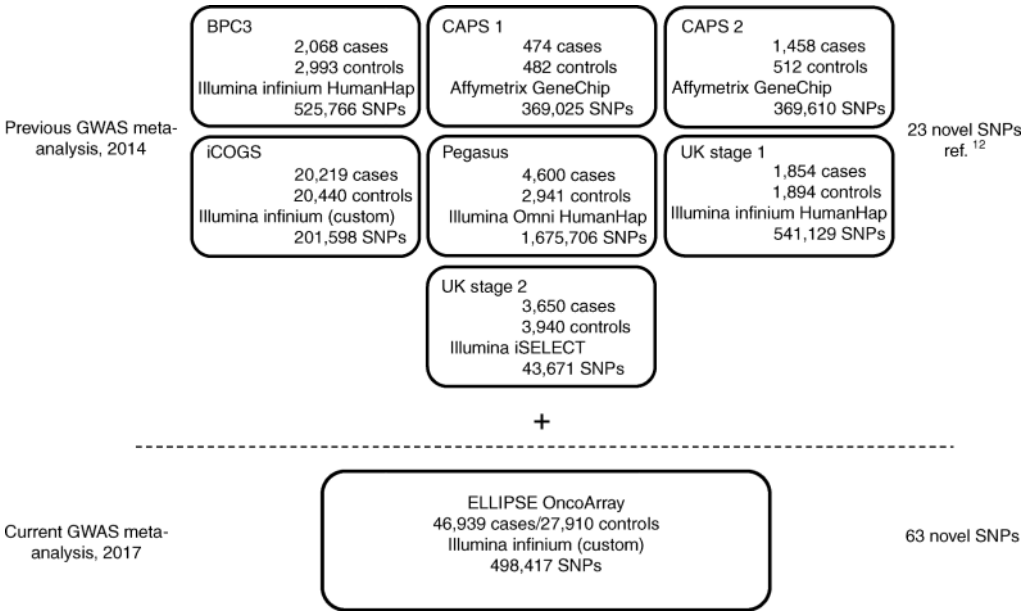

Fig. 1 |. ELLIPSE/PRACTICAL study overview of PrCa GWAS meta-analysis.

The top section describes the PrCa GWAS meta-analysis published in 2014, in which 23 novel variants were identified[12]. The current PrCa GWAS meta-analysis incorporated an additional 46,939 PrCa cases and 27,910 controls independent of the meta-analyses. The current meta-analysis discovered 62 novel variants associated with overall PrCa and 1 novel variant associated with early-onset PrCa.

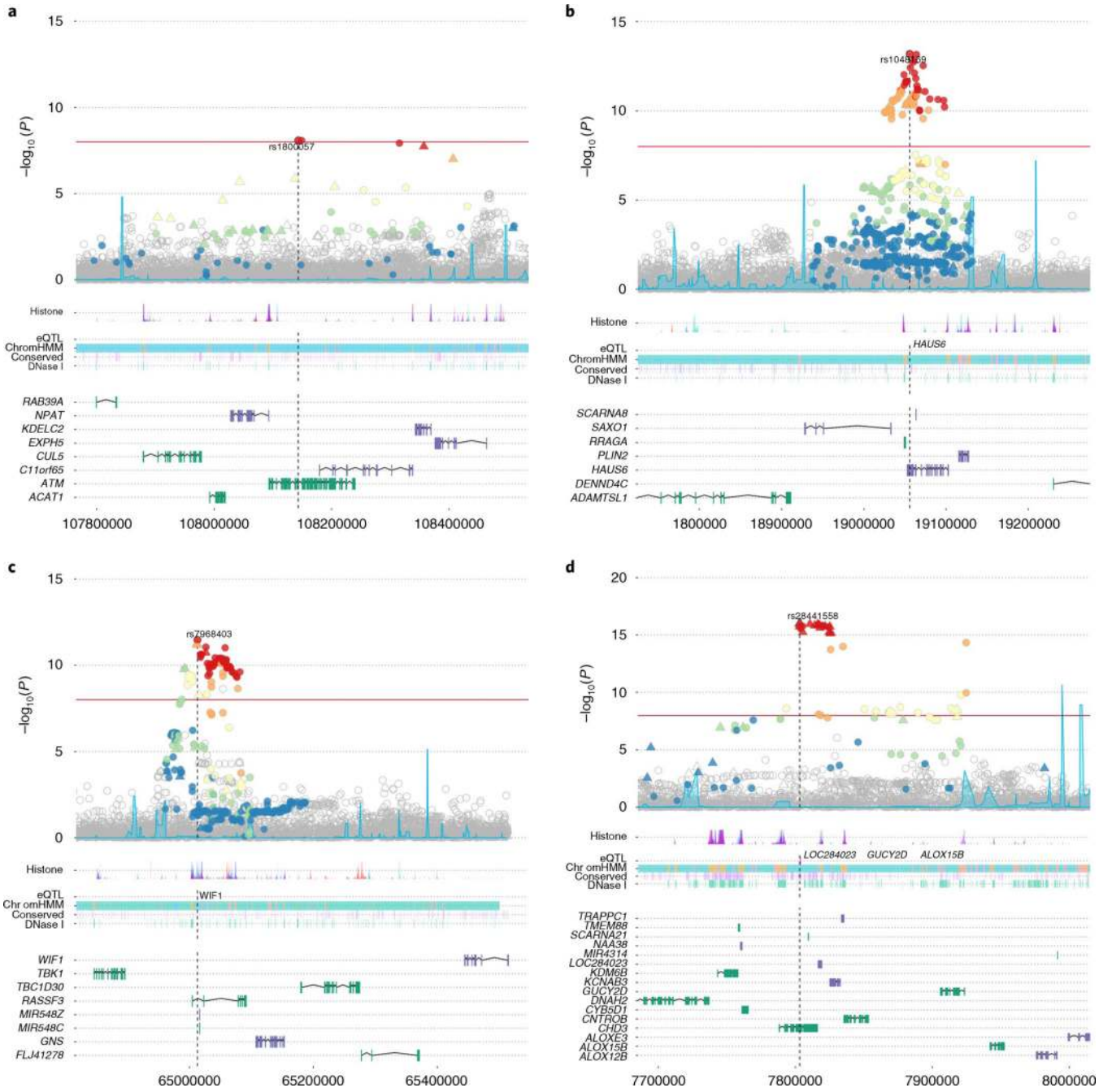

Fig. 2 |. Locus Explorer plots depicting the statistical association with PrCa and biological context of variants from four of the newly identified PrCa-risk loci (*n* = 74,849 biologically independent samples).

a-d, Top, Manhattan plots of variant $-\log_{10} P$ values (y axis), with the Index SNP labeled. Variants that were directly genotyped with the OncoArray are represented as triangles, and imputed variants are represented as circles. Variants in LD with the index SNP are denoted by color (red, $r^2 > 0.8$; orange, $0.6 < r^2 < 0.8$; yellow, $0.4 < r^2 < 0.6$; green, $0.2 < r^2 < 0.4$, blue, $r^2 \leq 0.2$). Middle, relative locations of selected biological annotations: histone marks within seven cell lines from the ENCODE project; genes for which the index SNP is an

eQTL in TCGA prostate adenocarcinoma dataset; chromatin state annotation by ChromHMM in PrEC cells; conserved elements within the genome; and DNAse I-hypersensitivity sites in ENCODE prostate cell lines. Bottom, positions of genes within the region, with genes on the positive and negative strands marked in green and purple, respectively. The horizontal axis represents genomic coordinates in the hg19 reference genome. **a**, rs1800057 (chromosome (chr) 11: 107643000–108644000). The index variant is a nonsynonymous SNP in *ATM*. **b**, rs1048160 (chr 9: 18556000–19557000). The index variant is located within the 3′ UTR of *HAUS6* and is an eQTL for *HAUS6.* **c**, rs7968403 (chr 12: 64513000–65514000). The signal is centered on *RASSF3,* and the index variant is located within the first intron. This SNP is also situated within a region annotated for multiple regulatory markers and is an eQTL for the more distant *WIF1* gene. **d**, rs28441558 (chr 17: 7303000–8304000). The signal implicates a cluster of highly correlated variants centered on *CHD3.* The index SNP is also an eQTL for three other more distantly located genes.

Table 1 |

Prostate cancer OncoArray and GWAS meta-analysis for 63 novel regions

SNP	Reference RAF[a]	Band	Position	Nearest gene	Alleles[b]	RAF	OR	95% CI	P[c]
Novel loci associated with overall prostate cancer									
rs56391074	0.329	1p22.3	88210715	*RP11–60A14.1*	AT/A	0.38	1.05	1.03–1.06	1.7×10^{-8}
rs34579442	0.316	1q21.3	153899900	*DENND4B*	C/CT	0.34	1.07	1.05–1.09	4.5×10^{-14}
rs62106670	0.400	2p25.1	8597123	*AC011747.3*	T/C	0.38	1.05	1.04–1.07	7.1×10^{-9}
rs74702681	0.024	2p14	66652885	*MEIS1-AS3*	T/C	0.02	1.17	1.11–1.23	2.0×10^{-9}
rs11691517	0.750	2q13	111893096	*BCL2L11*	T/G	0.74	1.07	1.05–1.08	3.5×10^{-12}
rs34925593	0.481	2q31.1	174234547	*CDCA7*	C/T	0.48	1.05	1.03–1.07	2.8×10^{-8}
rs59308963	0.726	2q33.1	202123479	*CASP8*	T/TATTCTGTC	0.73	1.05	1.03–1.07	2.4×10^{-8}
rs1283104	0.407	3q13.12	106962521	*DUBR*	G/C	0.38	1.05	1.03–1.07	8.8×10^{-9}
rs182314334	0.888	3q25.1	152004202	*MBNL1*	T/C	0.90	1.09	1.06–1.12	4.1×10^{-11}
rs142436749	0.012	3q26.2	169093100	*MECOM*	G/A	0.01	1.25	1.16–1.34	4.7×10^{-9}
rs10793821	0.580	5q31.1	133836209	*RNU6–456P*	T/C	0.57	1.05	1.04–1.07	5.4×10^{-11}
rs76551843	0.991	5q35.1	169172133	*DOCK2*	A/G	0.99	1.31	1.19–1.44	1.7×10^{-8}
rs4976790	0.096	5q35.3	177968915	*COL23A1*	T/G	0.11	1.08	1.05–1.10	6.7×10^{-9}
rs12665339	0.148	6p21.33	30601232	*ATAT1*	G/A	0.17	1.06	1.04–1.08	5.6×10^{-9}
rs9296068	0.645	6p21.32	32988695	*HLA-DOA*	T/G	0.65	1.05	1.03–1.07	1.3×10^{-8}
rs9469899	0.356	6p21.31	34793124	*UHRF1BP1*	A/G	0.36	1.05	1.03–1.07	5.3×10^{-9}
rs4711748	0.232	6p21.1	43694598	*RP1–261G23.5*	T/C	0.23	1.05	1.03–1.07	3.4×10^{-8}
rs527510716	0.251	7p22.3	1944537	*MAD1L1*	C/G	0.24	1.06	1.04–1.08	4.9×10^{-8}
rs11452686	0.567	7p21.1	20414110	*ITGB8*	T/TA	0.56	1.05	1.03–1.07	7.8×10^{-9}
rs17621345	0.758	7p14.1	40875192	*SUGCT*	A/C	0.74	1.07	1.05–1.09	6.7×10^{-14}
rs1048169	0.367	9p22.1	19055965	*HAUS6*	C/T	0.38	1.06	1.05–1.08	6.5×10^{-14}
rs10122495	0.296	9p13.3	34049779	*RN7SKP114*	T/A	0.31	1.05	1.03–1.07	1.3×10^{-8}
rs1182	0.258	9q34.11	132576060	*TOR1A*	A/C	0.22	1.06	1.04–1.08	1.1×10^{-9}
rs141536087	0.166	10p15.3	854691	*LARP4B*	GCGCA/G	0.15	1.08	1.06–1.11	9.0×10^{-13}
rs1935581	0.605	10q23.31	90195149	*RNLS*	C/T	0.63	1.05	1.03–1.07	6.5×10^{-9}
rs7094871	0.540	10q25.2	114712154	*TCF7L2*	G/C	0.54	1.04	1.03–1.06	4.8×10^{-8}

SNP	Reference RAF[a]	Band	Position	Nearest gene	Alleles[b]	RAF	OR	95% CI	P[c]
rs11666569	0.728	19p13.11	17214073	*MYO9B*	C/T	0.71	1.05	1.03–1.07	8.2×10^{-9}
rs118005503	0.912	19q12	32167803	*THEG5*	G/C	0.91	1.09	1.06–1.13	7.3×10^{-9}
rs61088131	0.848	19q13.2	42700947	*POU2F2*	T/C	0.82	1.06	1.04–1.09	8.8×10^{-9}
rs11480453	0.641	20q11.21	31347512	*DNMT3B*	C/CA	0.60	1.05	1.03–1.06	3.2×10^{-8}
rs6091758	0.465	20q13.2	52455205	*BCAS1*	G/A	0.47	1.07	1.06–1.09	6.4×10^{-18}
rs9625483	0.026	22q12.1	28888939	*TTC28*	A/G	0.03	1.14	1.09–1.20	2.4×10^{-8}
rs17321482	0.873	23p22.2	11482634	*ARHGAP6*	C/T	0.87	1.07	1.05–1.09	2.1×10^{-13}
Novel locus associated with early-onset prostate cancer									
rs138004030	0.920	6q27	170475879	*LOC154449*	G/A	0.91	1.27	1.17–1.38	2.9×10^{-8}

[a] Risk-allele frequency (RAF) in 1KGP Europeans.

[b] Risk allele/reference allele.

[c] *P* values generated from likelihood-ratio tests.

[d] Region previously reported by Wang et al.[49], rs12791447; rs61890184-rs12791447 r^2 (EUR) = 0.41.

Table 2 |

Estimation of polygenic risk scores by using 147 prostate cancer-susceptibility variants

Risk category percentile[a]	Relative risk	95% CI
<1	0.15	0.11–0.20
1–10	0.35	0.32–0.37
10–25	0.54	0.51–0.57
25–75	1.00 (baseline)	
75–90	1.74	1.67–1.82
90–99	2.69	2.55–2.82
≥99	5.71	5.04–6.48

[a]PRS percentiles based on the cumulative score distributed among controls. The beta coefficients computed from the European overall meta-analysis were applied to determine the PRS risk among individuals in the OncoArray study.